Supplemental Materials for

DNA methylation modulates transcription factor occupancy chiefly at sites of high intrinsic cell-type variability

Matthew T. Maurano, Hao Wang, Sam John, Anthony Shafer,
Theresa Canfield, Kristen Lee, John A. Stamatoyannopoulos

SUPPLEMENTAL FIGURES

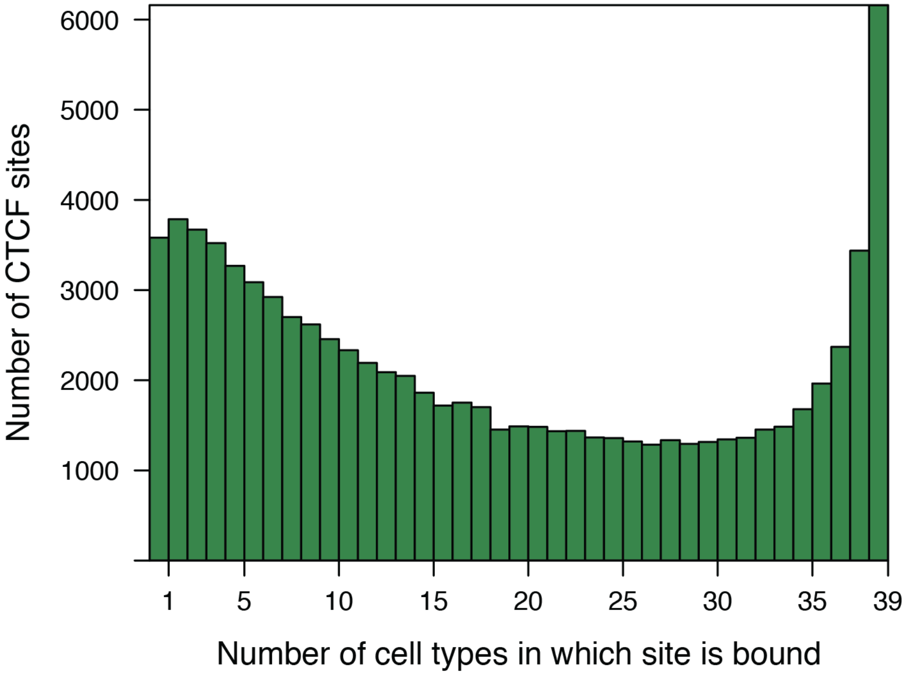

Fig. S1. Genome-wide analysis of CTCF binding in 40 cell types.
Variable sites exhibit a wide degree of cell-type selectivity across 40 cell types.

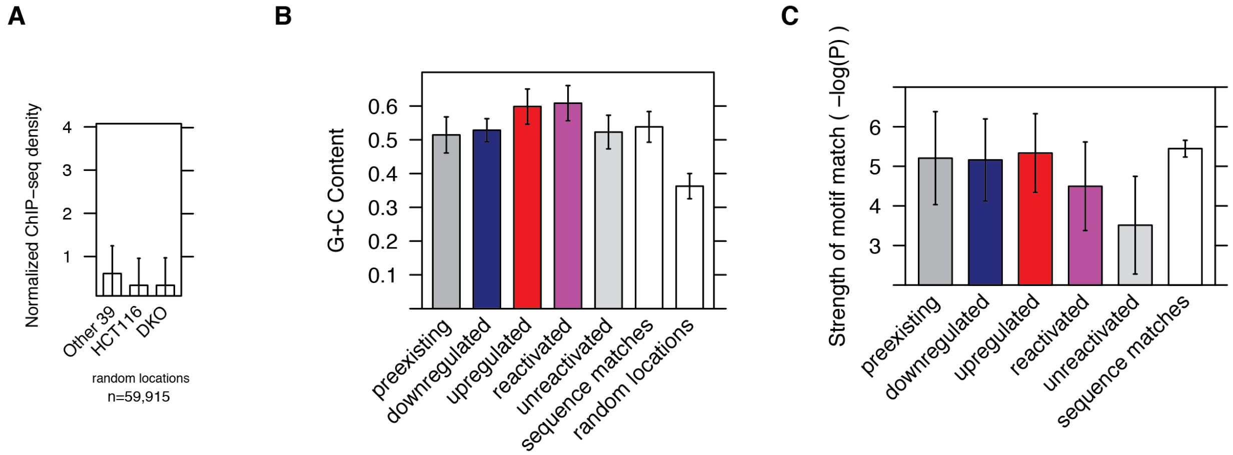

Fig. S2. Sequence characteristics of CTCF sites.

(A) Random genomic locations not matching the CTCF motif demonstrate similar occupancy to informatically predicted sequence matches to the CTCF binding motif (Fig. 1E).

(B) Strength of match to the consensus binding sequence by class of site. Note that sequence matches have as good or better matches when compared to in vivo determined binding sites.

(C) G+C content by class of site. Note increased G+C content for upregulated and reactivated sites. Note the selection of random locations without any match to the CTCF recognition sequence results in low G+C content.

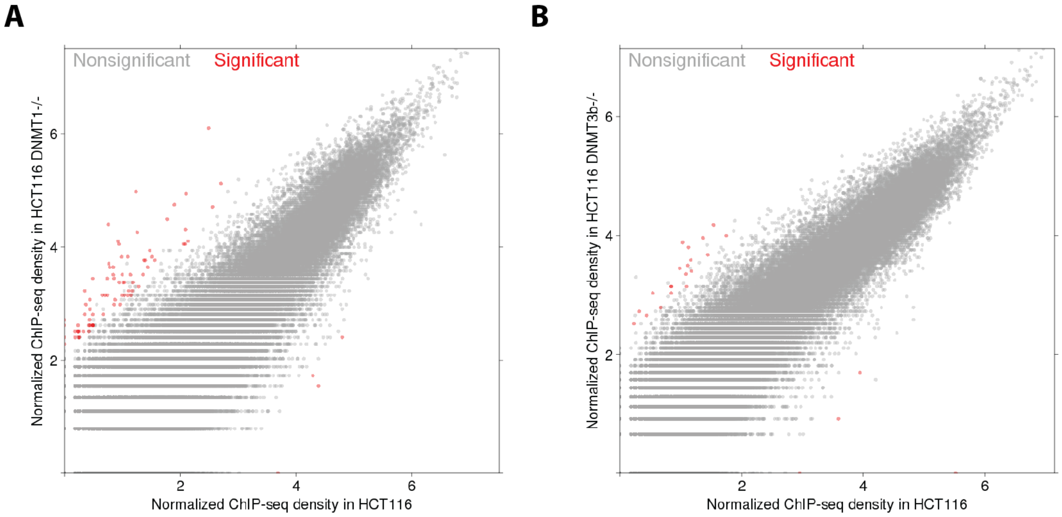

Fig. S3. Lack of CTCF reactivation in HCT116 DNMT1$^{-/-}$ (A) and DNMT3B (B)$^{-/-}$ cells.
Red indicates significant difference in CTCF occupancy by ChIP-seq. DNMT3B$^{-/-}$ cells demonstrated 20 upregulated sites and 4 downregulated sites, while DNMT1$^{-/-}$cells demonstrated 77 upregulated sites and 4 downregulated sites at 1% FDR.

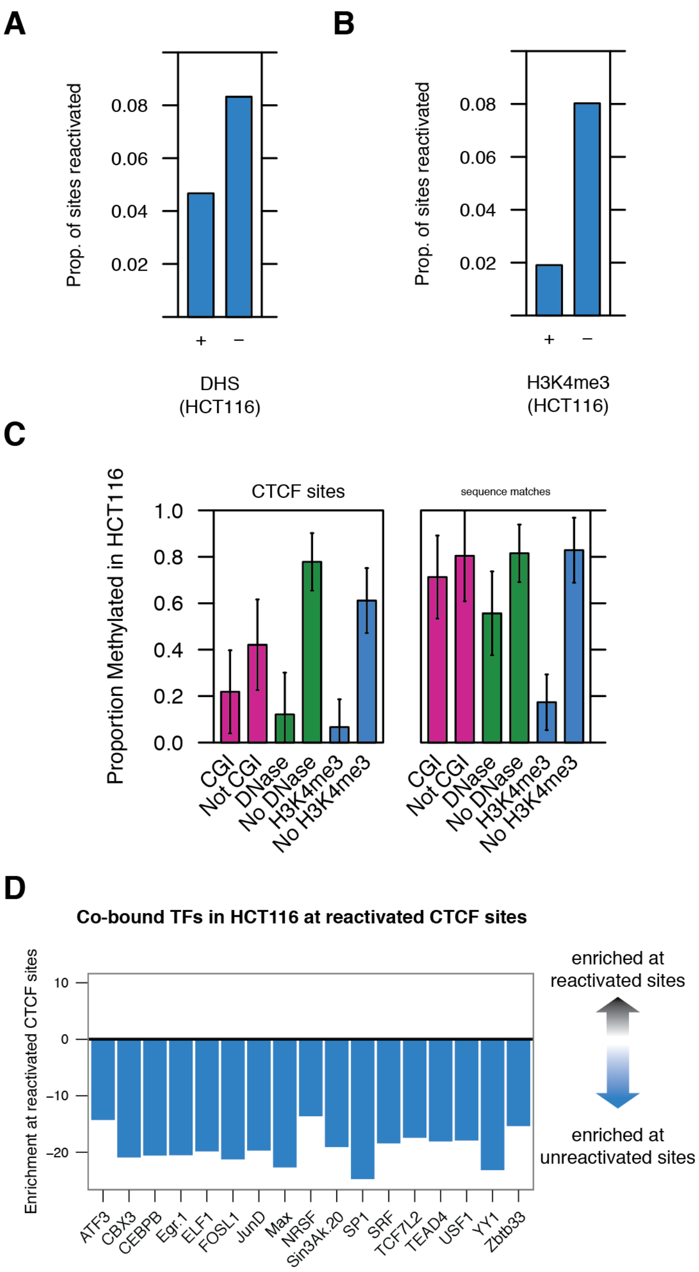

Fig. S4. Chromatin characteristics of response to demethylation.

(A-B) Reactivation frequency (reactivated sites as a proportion of both reactivated and reactivated sites) of CTCF sites outside or inside a DHS (A) or H3K4me3 peak (B).

(C) Methylation levels at sites grouped by CTCF reactivation (panels) and genomic features (x axis). The presence of a CpG island, DHSs or H3K4me3 peak is associated with reduced levels of methylation, though the relationship is strongest for DNase I and H3K4me3.

(D) Reactivated CTCF sites are depleted for co-occupancy of common TFs.

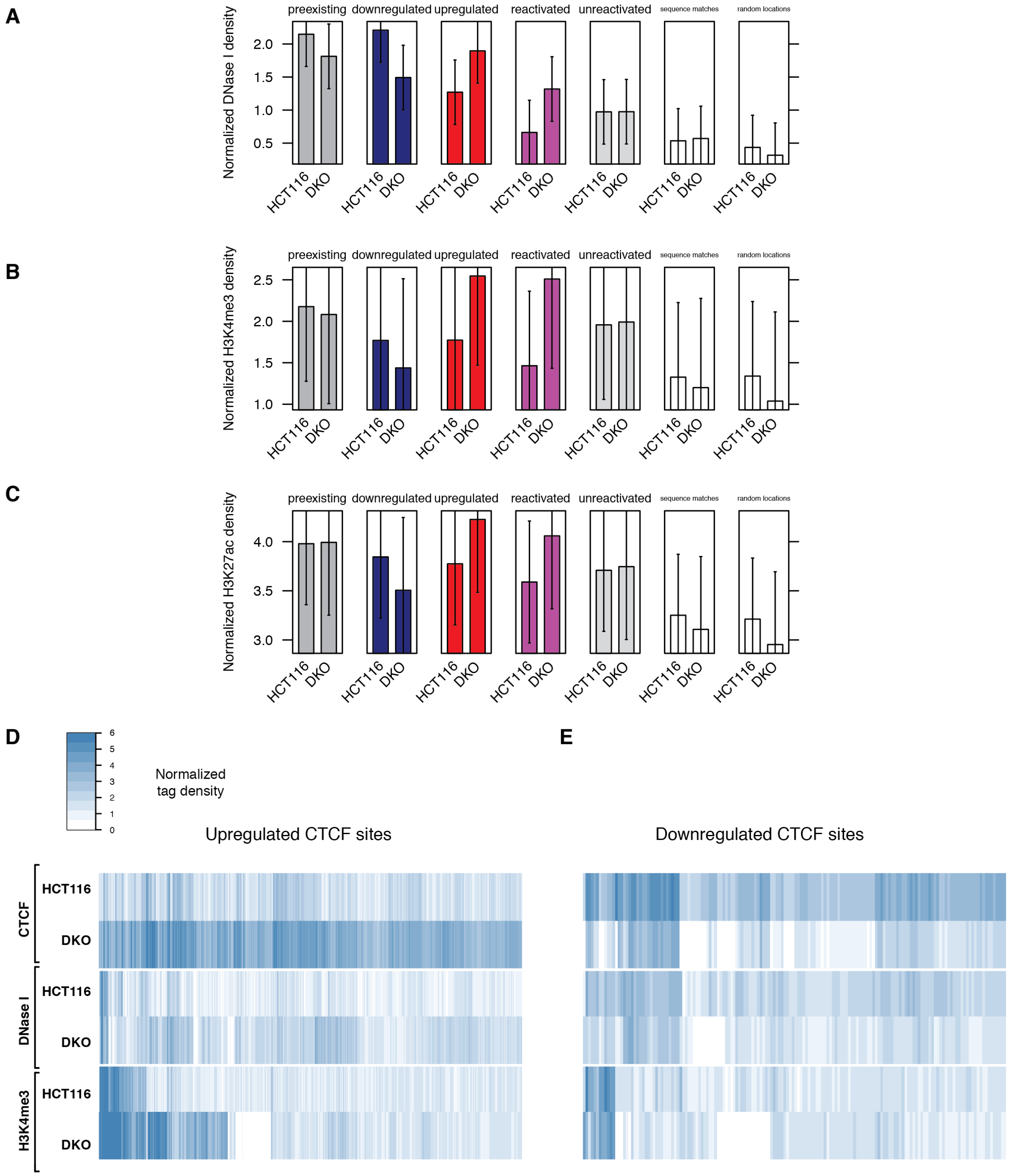

Fig. S5. Chromatin dynamics at demethylated CTCF sites.

(A-C) Quantitative comparison of mean DNase I (A), H3K4me3 (B) and H3K27ac (C) density across each set of sites in HCT116 and DKO cells.

(D-E) Normalized CTCF, DNaseI and H3K4me3 signal at sites upregulated (D) or downregulated (E) for CTCF signal.

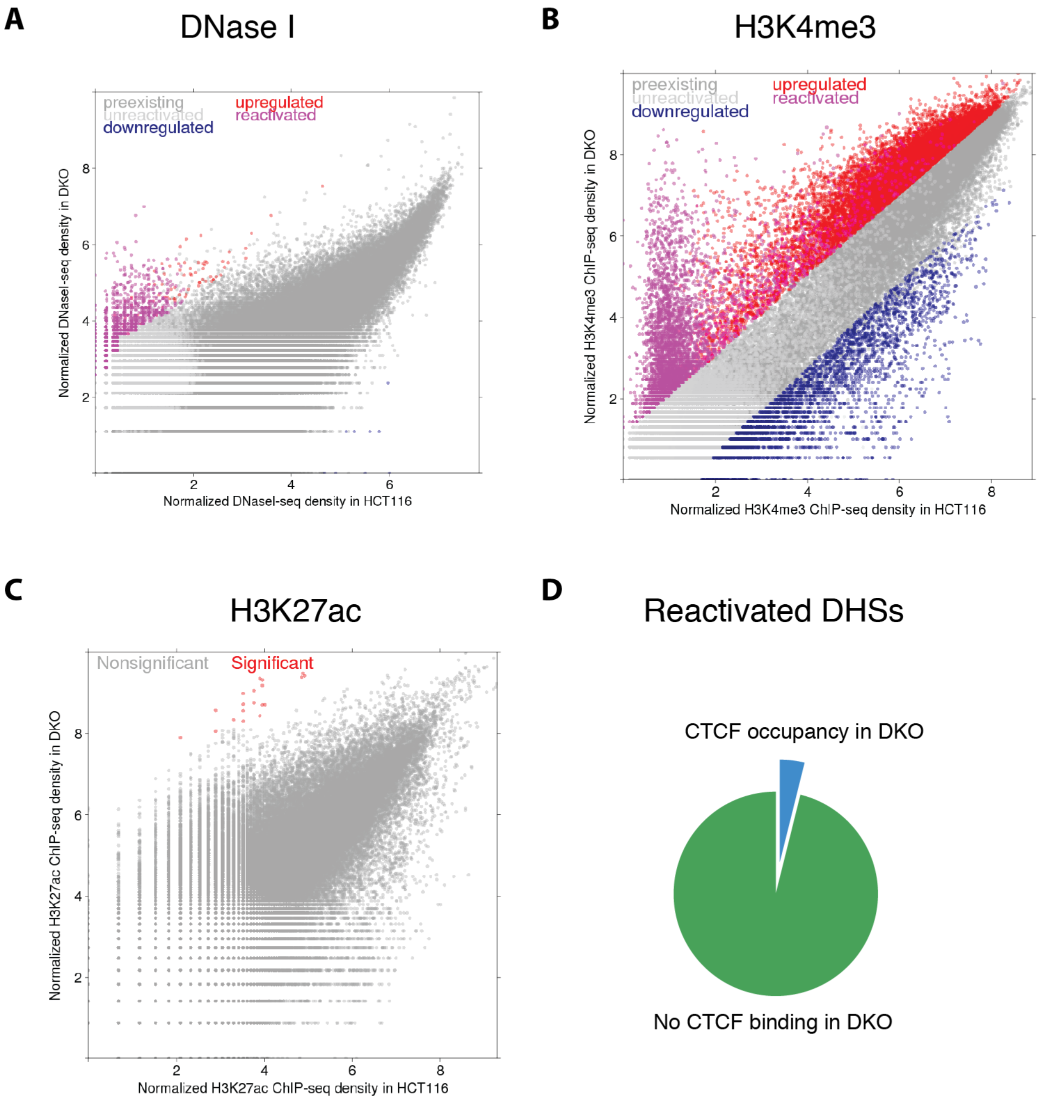

Fig. S6. Global reactivation of chromatin structure.

(A-C) Comparison of genome-wide DNaseI **(A)**, H3K4me3 **(B)** and H3K27ac **(C)** density between wild type HCT116 and DKO at DHSs. Red indicates significant difference in density.

(D) Broad extent of chromatin reactivation beyond CTCF sites.

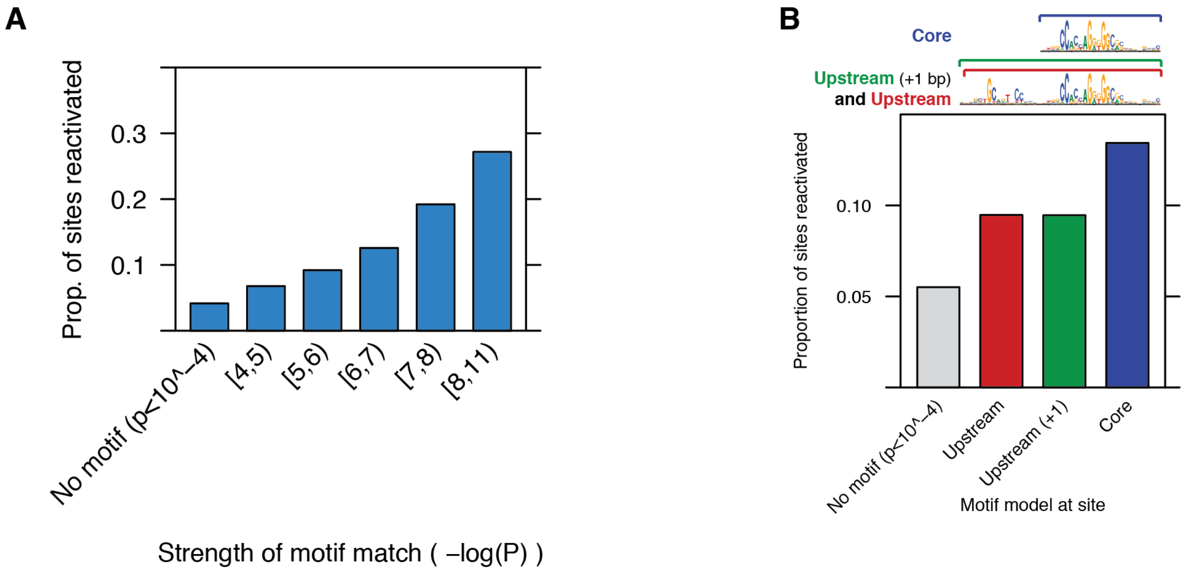

Fig. S7. Reactivation frequency relates to sequence characteristics of the CTCF recognition sequence. Reactivation frequency by strength of motif match (A) binding mode (B).

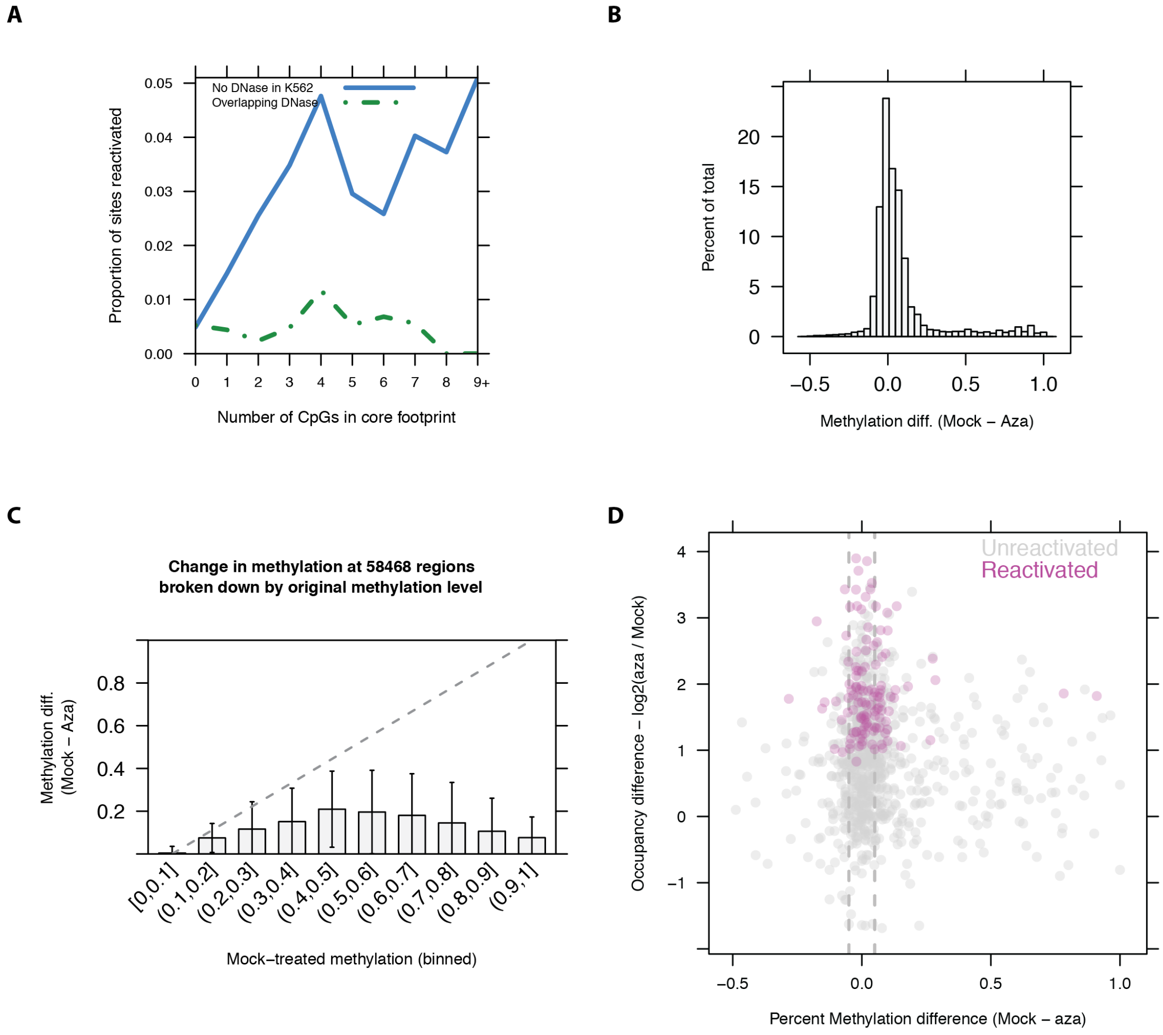

Fig. S8. Transient depletion of DNA methylation using 5-aza-2'-deoxycytidine.
(A) Reactivation frequency parallels local CpG content.

(B) Difference in methylation upon treatment with 5-aza-CdR at 11456 regions of preexisting (>50%) methylation in K562.

(C) Change in methylation upon 5-aza-CdR treatment for sites with varying levels of existing methylation in K562. Reduction of methylation is nearly complete at moderately methylated CpGs (<50%), but absolute change in methylation is reduced for sites demonstrating >50% methylation.

(D) Comparison between occupancy differences and methylation differences. Dashed vertical lines represent ±0.05 difference in proportion of methylation.

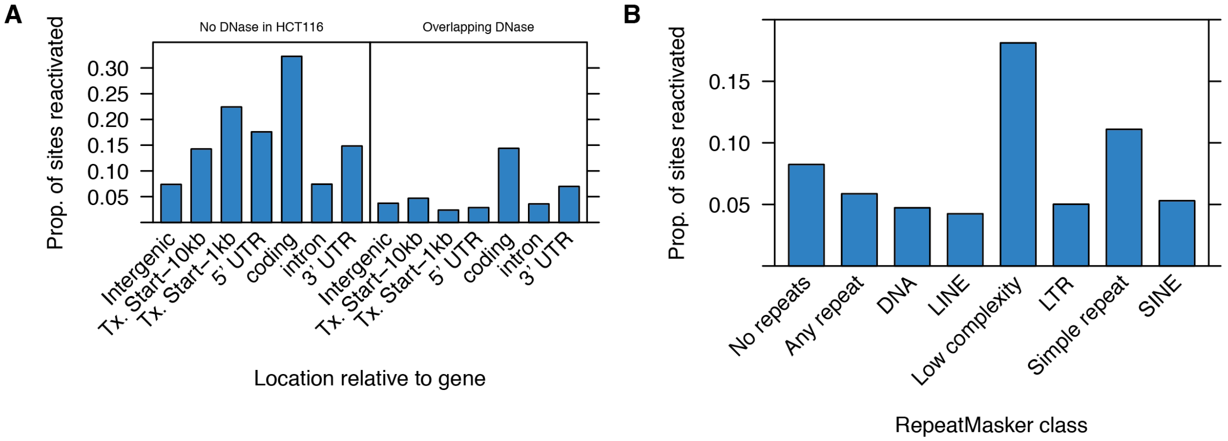

Fig. S9. Genomic characteristics associated with reactivation.
Reactivation frequency by genic location **(A)**, repeat class **(B)**.

SUPPLEMENTAL TABLES

Table S1. Growth conditions for new cell types.

Cell type	Description	Medium	Supplement				Notes
			%FBS	NEAA[a]	NaPyruvate[b]	Pen-strep[c]	
A549	human lung carcinoma epithelial cells	F-12K[d]	10%			1 ×	
AG04449	fetal buttock/thigh dermal fibroblast	MEM[e]	15%	1 ×		1 ×	
AG04450	fetal lung fibroblast	MEM[e]	15%	1 ×		1 ×	
BE2_C	human brain neuroblastoma cells	1:1 mixutre of F-12Kd and MEM[e]	10%	1 ×	1 ×	1 ×	
GM12878	human lymphoblastoid cell line	RPMI[f]	15%			1 ×	
HA-c	human astrocytes-cerebellar	AM astrocyte medium (ScienCell, #1801)					(h)
HA-sp	human astrocytes-spinal cord	AM astrocyte medium (ScienCell, #1801)					(h)
HCFaa	human cardiac fibroblasts-adult atrial	FM fibroblast medium (ScienCell, #2301)					(h)
HCM	human cardiac myocytes	CMM cardiac myocyte medium (ScienCell, #6201)					(h)
HCPEpiC	human choroid plexus epithelial cells	EpiCM, epithelial cell medium (ScienCell #4101)					(h)
HCT116	human colorectal carcinoma cells	McCoy's 5A modified medium (ATCC #30-2007)	10%			1 ×	
HCT116 DNMT1	human colorectal carcinoma cells with deletion of DNMT1	McCoy's 5A modified medium (ATCC #30-2007)	10%			1 ×	
HCT116 DNMT3B	human colorectal carcinoma cells with deletion of DNMT3b	McCoy's 5A modified medium (ATCC #30-2007)	10%			1 ×	
HCT116 DKO	human colorectal carcinoma cells with deletion of DNMT1 and DNMT3b	McCoy's 5A modified medium (ATCC #30-2007)	10%			1 ×	
HEK-293	human embryonic kidney epithelial cells	MEM[e]	10%	1 ×	1 ×	1 ×	
HFF	human foreskin fibroblasts	DMEM[g]	10%			1 ×	
HFF-Myc	human foreskin fibroblasts expressing canine cMyc	DMEM[g]	10%			1 ×	
HMEC	human mammary epithelial cells	MEGM BulletKit (Lonza #CC-3150)					
HRPEpiC	human retinal pigment epithelial cells	EpiCM, epithelial cell medium (ScienCell #4101)					(h)
HUVEC	human umbilical vein endothelial cells	EGM-2 BulletKit (Lonza #CC-3162)					
HVMF	human villous mesenchymal fibroblast cells	FM fibroblast medium (ScienCell, #2301)					(h)
K562 aza	human chronic myelogenous leukemia cells treated with 1 μM 5-aza-2'-deoxycytidine in DMSO daily for 3 days	RPMI[f]	10%		1 ×	1 x	
K562 mock	human chronic myelogenous leukemia cells treated with 1 μM DMSO daily for 3 days	RPMI[f]	10%		1 ×	1 x	
MCF7	human mammary gland adenocarcinoma	MEM[e]	10%	1 ×		1 ×	bovine insulin[i]
NB4	human acute promyelocytic leukemia cells	RPMI[f]	10%			1 ×	
NHDF-Neo	neonatal normal human dermal fibroblasts	FGM-2 BulletKit (Lonza #CC-3132)					
NHEK	normal human epidermal keratinocytes	KGM BulletKit (Lonza #CC-3111)					
NHLF	normal human lung fibroblasts	FGM-2 Bullet Kit (Lonza #CC-3132)					
RPTEC	renal proximal tubule epithelial cells	REGM™BulletKit® (Lonza #CC-3190)					
WI-38/hTERT/GFP-RAF-ER	human embryonic lung fibroblasts	MEM[e]	10%	1 ×	1 ×	1 ×	

Supplements:
a: NEAA 0.1 mM non-essential amino acids (Invitrogen, #11140-050)
b: NaPyruvate 1 mM sodium pyruvate (Cellgro, #25-000-CI)
c: Pen-Strep 100 IU penicillin and 100 μg/ml streptomycin (Cellgro, #30-002CI)

Media:
d: F-12K Kaighn's modification of Ham's F-12 medium (ATCC, #30-2004)
e: MEM MEM with 2mM L-glutamine and Earle's salts (Cellgro, #10-010-CM)
f: RPMI RPMI 1640 with 2 mM L-glutamine (Cellgro, #10-040-CM)
g: DMEM DMEM with 4.5 g/L glucose, L-glutamine, and sodium pyruvate (Cellgro, 10-013-CV)

Notes:
h: coat flasks with poly-L-lysine2 μg/cm2, ScienCell #0413
i: bovine insulin with 0.01 mg/ml bovine insulin (Sigma #I6634)

Other cells cultured as previously described (Wang 2012).

Table S2. Summary of CTCF ChIP-seq experiments.

Cell type	# of non-redundant tags	SPOT Rep1	SPOT Rep2	SPOT Rep3	Published	Accession
A549	14061527	0.4586	0.4349		Novel	GSE30263
AG04449	29565255	0.6157	0.2732		Novel	GSE30263
AG04450	35133130	0.2939	0.3082		Novel	GSE30263
AG09309	22121486	0.3901	0.264		Wang 2012	GSE30263
AG09319	39296604	0.4307	0.3251		Wang 2012	GSE30263
AG10803	48734285	0.3229	0.2775		Wang 2012	GSE30263
AoAF	34080709	0.3891	0.3219		Wang 2012	GSE30263
BE2_C	38938541	0.3717	0.3043		Novel	GSE30263
BJ	20154780	0.5473	0.4401		Wang 2012	GSE30263
Caco_2	22633608	0.4291	0.2837		Wang 2012	GSE30263
GM06990	17564629	0.4851	0.446		Wang 2012	GSE30263
GM12878	19651161	0.445	0.3454		Novel	GSE30263
HAc	30983231	0.3286	0.2379		Novel	GSE30263
HAsp	26294223	0.3498	0.2369		Novel	GSE30263
HBMEC	31277391	0.542	0.2946		Wang 2012	GSE30263
HCM	27614074	0.4139	0.3096		Novel	GSE30263
HCPEpiC	41581449	0.6368	0.3573		Novel	GSE30263
HEEpiC	29763677	0.3467	0.3198		Wang 2012	GSE30263
HEK293	15353682	0.5978	0.4668		Novel	GSE30263
HeLa_S3	13695593	0.4769	0.4253		Wang 2012	GSE30263
HepG2	18174841	0.499	0.4195		Wang 2012	GSE30263
HFF_Myc	31402050	0.2936	0.2287		Novel	GSE30263
HMEC	27836859	0.566	0.3643		Novel	GSE30263
HMF	27854938	0.414	0.4151		Wang 2012	GSE30263
HPAF	46244156	0.4154	0.3409		Wang 2012	GSE30263
HPF	41607137	0.3115	0.2477		Wang 2012	GSE30263
HRE	12267228	0.5636	0.5548		Wang 2012	GSE30263
HRPEpiC	36393999	0.3228	0.3102		Novel	GSE30263
HUVEC	10749551	0.6721	0.5917		Novel	GSE30263
HVMF	22608466	0.3393	0.3958		Novel	GSE30263
K562	20398176	0.4664	0.4433		Wang 2012	GSE30263
MCF_7	38770153	0.3355	0.2505		Novel	GSE30263
NHDF_Neo	35771540	0.3492	0.2522		Novel	GSE30263
NHEK	12262582	0.6636	0.453		Novel	GSE30263
RPTEC	31992786	0.4334	0.3625		Novel	GSE30263
SAEC	13092335	0.6743	0.645		Wang 2012	GSE30263
Sk_N_Sh_RA	16616235	0.5992	0.6092		Wang 2012	GSE30263
WERI_Rb_1	18427416	0.571	0.4352		Wang 2012	GSE30263
WI_38	21034423	0.2201	0.2104		Novel	GSE30263
K562-Mock	38165687	0.2677	0.1259	0.3027	Novel	GSE50610
K562-aza	25113857	0.2298	0.2008	0.2111	Novel	GSE50610
HCT116	32270754	0.2999	0.2362	0.2531	Novel	GSE50610
DNMT1KO	9645811	0.196			Novel	GSE50610
DNMT3bKO	12618070	0.2367			Novel	GSE50610
DKO	16167225	0.2351	0.2393		Novel	GSE50610

Number of tags represents the total uniquely-mapped reads (excluding PCR duplicates) for all biological replicates of each cell type. Number of peaks refers to the number of CTCF binding sites considered present in a given cell type (see Methods). Pearson correlation coefficient (R) measures concordance between replicate normalized density tracks at peaks. Peak concordance between replicates refers to the average of the proportion of peaks in one replicate that are found in the second replicate and vice versa. 19 cell types have been previously published (Wang 2012). SPOT score indicates data quality.

Table S3. Sites susceptible to reactivation.

CTCF class	HCT116 mean proportion of methylation	# sites >25% methylation in HCT116	# monitored sites	# sites with CpG in core 44 bp	# sites in CpG islands
preexisting	0.12	680	5065	4807	3190
downregulated	0.36	3	7	6	1
upregulated	0.55	96	138	137	89
reactivated	0.85	647	675	670	439
unreactivated	0.44	1688	3323	3129	2111
sequence matches	0.78	7460	8293	7417	2101
random locations	0.82	5	6	4	0

Methylation levels and CpG content at potential CTCF binding sites. Shown are sites for which methylation data are available in HCT116.

Table S4. Summary of H3K4me3 and H3K27ac ChIP-seq experiments.

Cell type	Experiment	DS#	# of non-redundant tags	#Peaks	SPOT Score	Published?	Accession
HCT116	H3K4me3	DS16055	20347444	25747	0.6379	Thurman et al	GSE50610 (originally released in GSE35583)
HCT116	H3K4me3	DS16056	17659687	26088	0.6978	Thurman et al	GSE50610 (originally released in GSE35583)
DKO	H3K4me3	DS22869	8848852	25567	0.7665	Novel	GSE50610
HCT116	H3K27ac	DS22861	10270899	24658	0.3023	Novel	GSE50610
DKO	H3K27ac	DS22870	7700567	40502	0.465	Novel	GSE50610

Experiment indicates the antibody used. Number of tags represents the total uniquely-mapped reads (excluding PCR duplicates) for all biological replicates of each cell type. Number of peaks refers to the number of peaks using hotspot. 2 K562 and all HCT116 H3K4me3 data has been published (Thurman 2012) and represents ENCODE clone. SPOT score indicates data quality.

Table S5. Summary of DNase I experiments.

Cell type	Experiment	# of non-redundant tags	#DHSs	SPOT Score	Published?	Accession
HCT116	DNaseI	23562622	114051	0.4056	Thurman et al	GSE50610 (originally released in GSE29692)
HCT116	DNaseI	19730809	113968	0.4544	Thurman et al	GSE50610 (originally released in GSE29692)
HCT116	DNaseI	26263533	104193	0.3274	Novel	GSE50610
DKO	DNaseI	14353151	174936	0.3294	Novel	GSE50610

Number of tags represents the total uniquely-mapped reads (excluding PCR duplicates) for all biological replicates of each cell type. #DHSs represents 1% FDR hotspots in HCT116, 0.5% hotspots in DKO. SPOT score indicates data quality.

Table S6. Qualitative chromatin dynamics at CTCF sites.

		H3K4me3		H3K27ac		DNase I	
CTCF class	**# sites**	**HCT116**	**DKO**	**HCT116**	**DKO**	**HCT116**	**DKO**
preexisting	59915	10219	7796	6144	6885	47755	42078
downregulated	158	13	2	7	2	148	100
upregulated	809	78	189	35	89	462	666
reactivated	3237	91	705	36	293	544	1789
unreactivated	43176	4752	4961	3111	4240	11106	13856
sequence matches	176630	1833	2769	1092	2333	13442	22754
random locations	7486	33	18	26	43	176	338

For each class the total number of CTCF sites is shown, in addition to the subset of sites with H3K4me3, H3K27ac and DNase I peaks in HCT116 and/or DKO.

Table S7. Summary of capture bisulfite experiments.

Sample	# of tags	# of tags on target	# of non-redundant tags	#CpGs covered >8x	#CTCF sites covered >8x	Published?	Accession
K562-Mock	153040607	101382521	75931601	1665046	19775	Novel	GSE50610
K562-aza	136469310	98742701	29584174	813891	12860	Novel	GSE50610

Number of tags refers to the total mapped tags; # of tags on target refers to the capture regions; # of non-redundant tags represents the total uniquely-mapped reads excluding PCR duplicates. Number of CTCF sites counts the number of overlapping at least one CpG with >8x read depth.

Table S8. ChIP-seq binding positions in 40 cell types, DKO, and aza-treated K562 cells.

109,342 peaks (106,196 excluding demethylated cell types). chrom, chromStart, chromEnd, ID: hg19 coordinates of CTCF peak. fimo.pval, Strand, recognition.sequence, MotifModel: CTCF binding mode based on local sequence. mcv (for multiple cell verified): the number of cell types in which the site is occupied.

This table is uploaded as a separate file.